Sitemap

Component for Joomla!

This manual documents **version 3.15.x** of the Joomla! extension.

http://www.aimy-extensions.com/joomla/sitemap.html

Contents

1 Introduction **3**

2 Sitemap Features **3**

3 Technical Requirements to Use Aimy Sitemap Generator **3**

4 Installing the Sitemap Extension **4**

5 Workflow **4**

6 Configuring the Sitemap Extension **4**
6.1 Aimy Sitemap: Preferences 5
6.1.1 Tab: Aimy Sitemap 5
6.1.2 Tab: Default Values 5
6.1.3 Tab: Crawler 6
6.1.4 Tab: Notifier 7
6.1.5 Tab: Periodic Crawl (PRO feature) 7
6.1.6 Tab: Permissions 8

7 The robots.txt File **9**
7.1 Editing 9
7.2 How To Set Rules For Aimy Sitemap's Crawler 9
7.3 Further Information 9

8 Crawling Your Joomla! Website **10**

9 Managing Your Index and Sitemap **10**
9.1 Setting Attributes 11
9.2 Searching, Filtering and Sorting 12
9.3 When to Re-Crawl Your Website? 13

10 Notifying Search Engines **14**

11 Generating an HTML Sitemap **14**
11.1 Options 15

12 Periodic Crawling (PRO feature) **17**
12.1 Obtaining the Script's Path 17
12.2 Scheduling Automatic Crawls 17
12.2.1 Setting Up a Cron-Job (Unix) 17
12.2.2 Setting Up a New Task (Windows) 18
12.2.3 Using Your Hosters Interface 19
12.3 Crawling Information 19

13 Special Cases **21**
13.1 Crawling HTTPS-only Websites (PRO feature) 21
13.2 Joomla! Installation in Subdirectory 21
13.2.1 Rewrite Rule 21
13.2.2 Sitemap Index File 21
13.2.3 Robots.txt Directive 22
13.3 Using a Canonical URL (PRO feature) 22

14 Notes and Limitations **22**

15 Debugging and Errors **23**

16 Copyright & Trademark Notice **23**

1 Introduction

The Joomla! component Aimy Sitemap generates an XML sitemap of your website for search engines. You may generate an HTML sitemap for your visitors as well.

You can customize the sitemap with different options and notify search engines about changes to your sitemap directly from the Joomla! backend.

Aimy Sitemap generator includes a crawler that visits every page of your website, analyzes its content and queues its URL for inclusion in your sitemap.

This manual guides you through all steps necessary to install, configure and use the extension to enrich your website and increase your SEO results.

2 Sitemap Features

Aimy Sitemap currently provides the following features:

- Integrated crawler that creates an index of your website
- Evaluation of `robots.txt` rules
- Customizable crawler delay and timeout settings
- Customizable exclude patterns
- Customizable *default* and *per URL* settings for priority, change frequency and inclusion in your sitemap
- Crawling of HTTPS-only websites (PRO feature)
- XML sitemap for search engines
- HTML sitemap for human visitors
  - List & Index view
  - Hierarchy view (PRO feature)
  - Change displayed titles (PRO feature)
  - Multi-language site support (PRO feature)
- Notification of search engines directly from your Joomla! backend
- Integrated `robots.txt` editor
- Script for periodic crawling via cron (Unix) / Task Scheduler (Windows) (PRO feature)
- Extension's interface languages: English, Spanish, German, Polish, Slovak, Farsi (Iran), Dutch, French, Portuguese (Brazil)
- Zero dependencies besides those of Joomla!, neither cURL nor url-fopen are required

3 Technical Requirements to Use Aimy Sitemap Generator

Aimy Sitemap supports **MySQL** and, in the PRO version, **PostgreSQL** databases and requires **PHP version 5.3.10 or higher** (like Joomla! 3 itself).

In order to crawl HTTPS-only websites, your PHP installation has to support SSL.

To automatically crawl your Joomla! website periodically, a system service like **cron** (Unix) or **Task Scheduler** (Windows) is required. Most hosters provide an interface to such services as well.

4 Installing the Sitemap Extension

The installation of the sitemap extension follows the common Joomla! procedures.

In case you are not familiar with these procedures proceed as follows:

1. Download the extension's ZIP archive
2. Log into your Joomla! backend as "Super User"
3. From the menu, choose "*Extensions*" → "*Extensions Manager*"
4. Click on the "*Browse*" button and select the ZIP archive
5. Click on the "*Upload & Install*" button

The extension's archive will be uploaded and installed afterwards.

For further information, please have a look at the Joomla! documentation "Installing an Extension".

5 Workflow

The common workflow for generating a sitemap is as follows:

1. Configure Aimy Sitemap
2. Edit your robots.txt file (optional)
3. Crawl your Joomla! website
4. View and customize the index of your sitemap
5. Write the sitemap to your webspace
6. Notify search engines (optional)

Whenever you want to update your sitemap repeat steps 3 to 6.

If you use Aimy Sitemap PRO you may update your sitemap automatically with *Periodic Crawls*, explained later on in this manual.

6 Configuring the Sitemap Extension

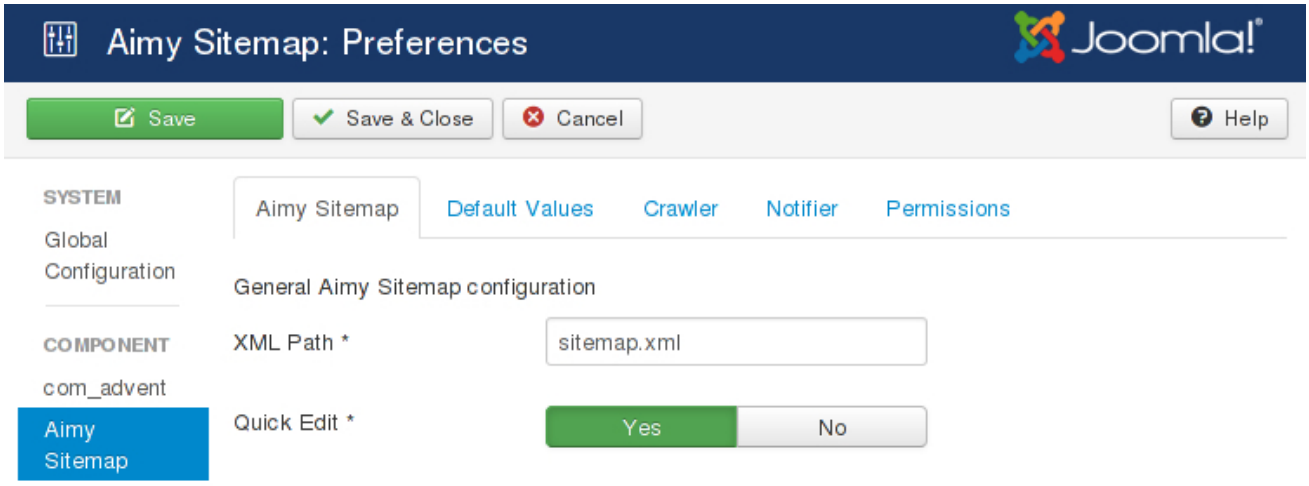
After a fresh installation, click on the "*Edit your sitemap now*" button on the installation report page.

At any time, you can configure the Aimy Sitemap extension using Joomla!'s Component Manager by choosing "*Components*" → "*Aimy Sitemap*" from the menu. To set your preferences, click the *Options* button.

6.1 Aimy Sitemap: Preferences

Aimy Sitemap allows you the customize sitemap, crawler, notifier and access permissions in a couple of ways. Let's have a look at the different tabs.

6.1.1 Tab: Aimy Sitemap



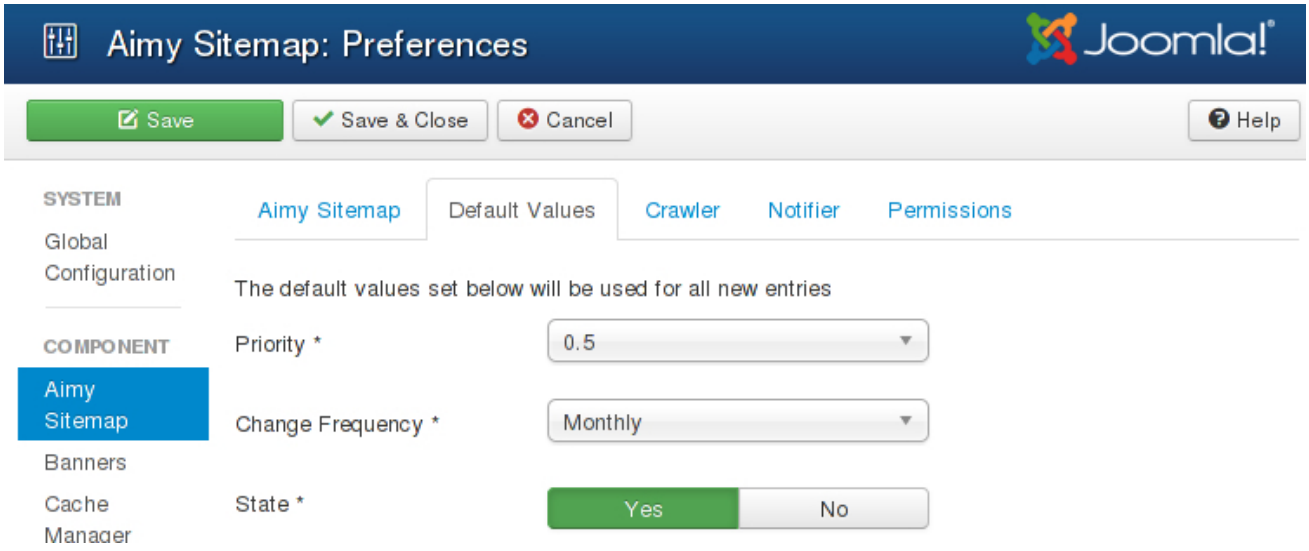
- **XML Path**: Specify a custom path for your sitemap in case you do not want to use the common *sitemap.xml* path.

  Any path is relative to the root directory of your Joomla! installation.

- **Quick Edit**: Enable or disable quick edit for managing URLs.

  If enabled, "*Manage URLs*" will allow you to edit your URLs' attributes directly from the list using drop-down boxes and buttons (using AJAX).

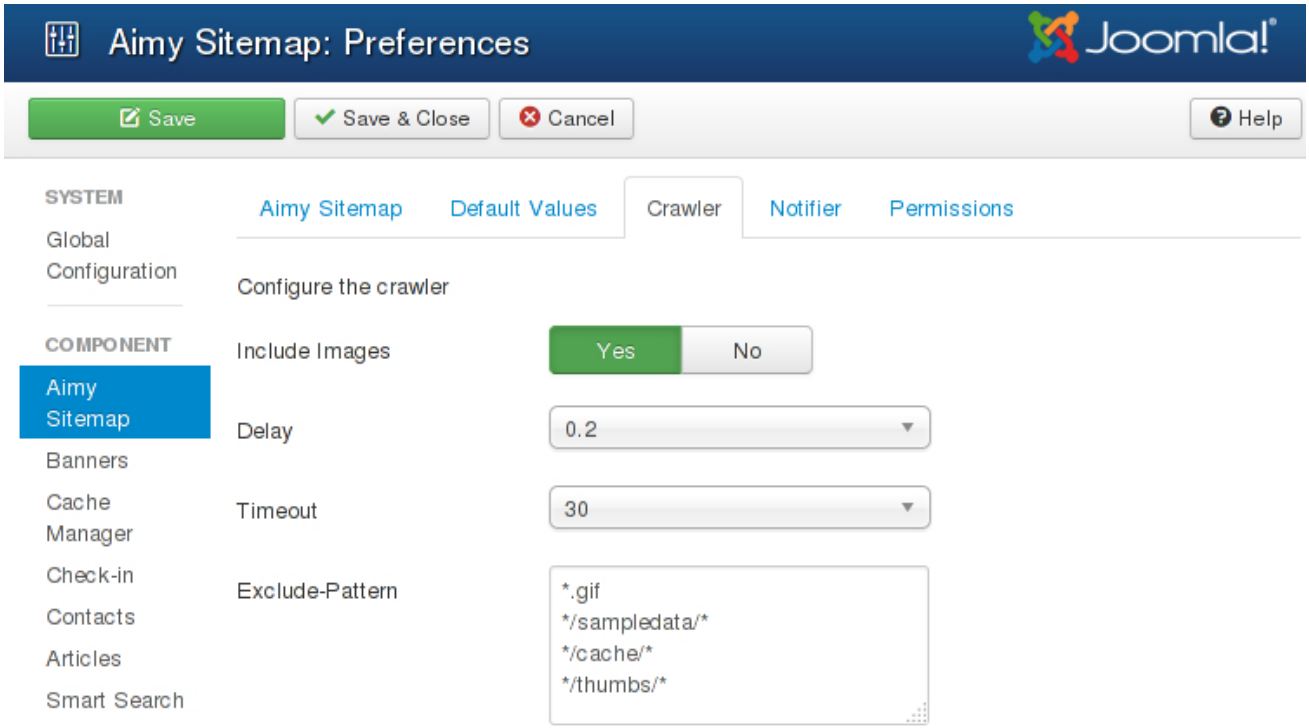
6.1.2 Tab: Default Values



- **Priority**: Set a default value for the priority of your documents and items. With the priority you can give search engines a hint on the importance of your pages. Valid values range from 0.0 to 1.0.

- **Change Frequency**: Set a default value for how often the documents of your website change.
- **State**: Define whether *new* documents should be added to the sitemap by default. You may deactivate single items individually, see *"Updating your Sitemap"*.

6.1.3 Tab: Crawler



- **Include Images**: Choose whether images should be added to your sitemap. If enabled, the crawler will add all found images that are referenced by the HTML tag `img` to your index. External sources not located on your website will not be added.
- **Delay**: The crawler goes through your Joomla! website one resource after another. You may set a delay in seconds. The crawler will wait between each request so your server is not put under too high load.
- **Timeout**: You can define a timeout in seconds that sets the time, the crawler will at maximum wait for your webserver to respond to a request. If the timeout is reached, the crawler will abort.

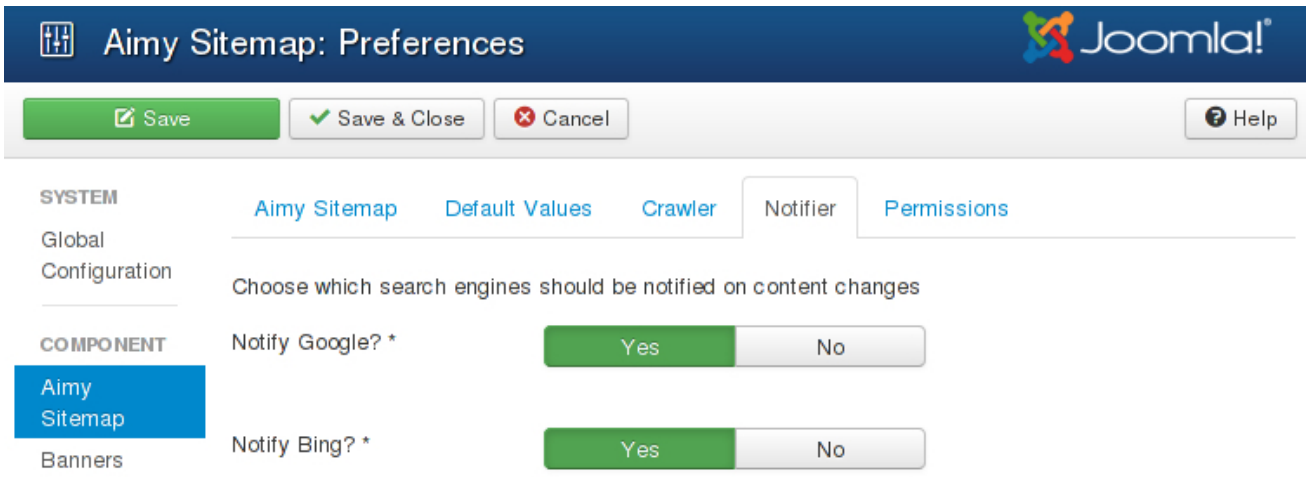
  If your webserver is rather slow, you may want to increase this value to at least 30 seconds.
- **Exclude-Pattern**: You may specify wildcard patterns, one on each line, that should be *excluded* from your sitemap. These patterns are case-sensitive. Regular expressions are not supported.

  Examples:

  - `*.gif` will exclude any file having the extension "`.gif`".
  - `*/sampledata/*` excludes any files that contain "`/sampledata/`" in their URL, like "`/images/sampledata/apple.jpg`".

Note: If you update the crawler's configuration at any later time, please crawl the website again to apply your changes.

6.1.4 Tab: Notifier

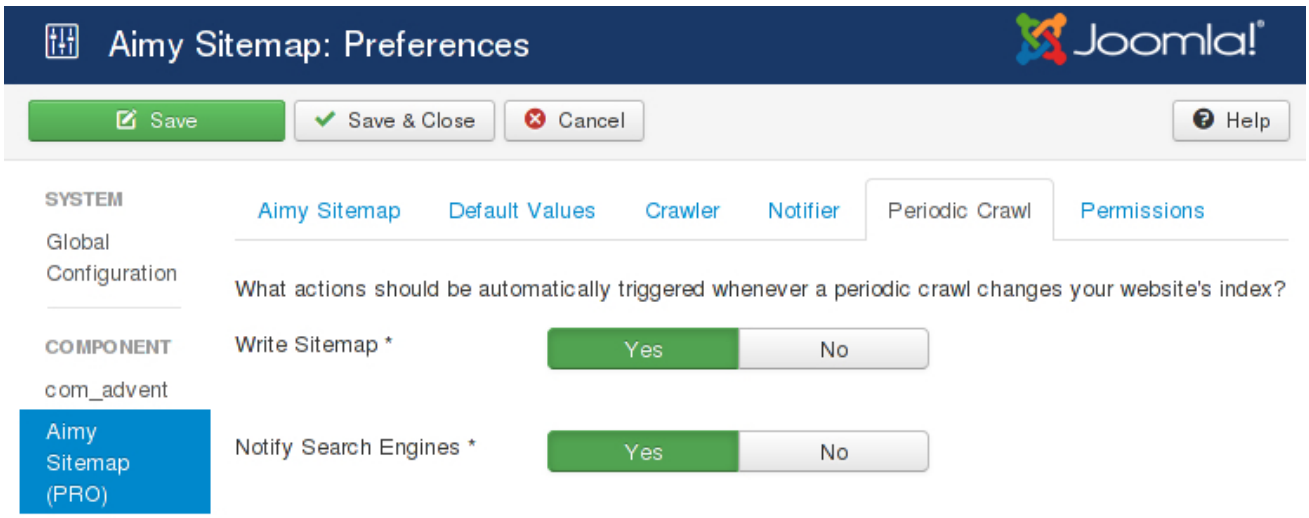


Choose which search engines should be notified by Aimy Sitemap's notifier. The two most widely used search engines, Google and Bing, are supported.

Note: The notification will not be done automatically. If you have generated or updated your sitemap, go to "*Components*" → "*Aimy Sitemap*" → "*Notify Search Engines*" and click the "*Start Notifying*" button to start the notification from the Joomla! backend.

6.1.5 Tab: Periodic Crawl (PRO feature)

Aimy Sitemap allows to set up actions that are triggered whenever a periodic crawl finishes and new or updated content has been found on your website.



Write Sitemap

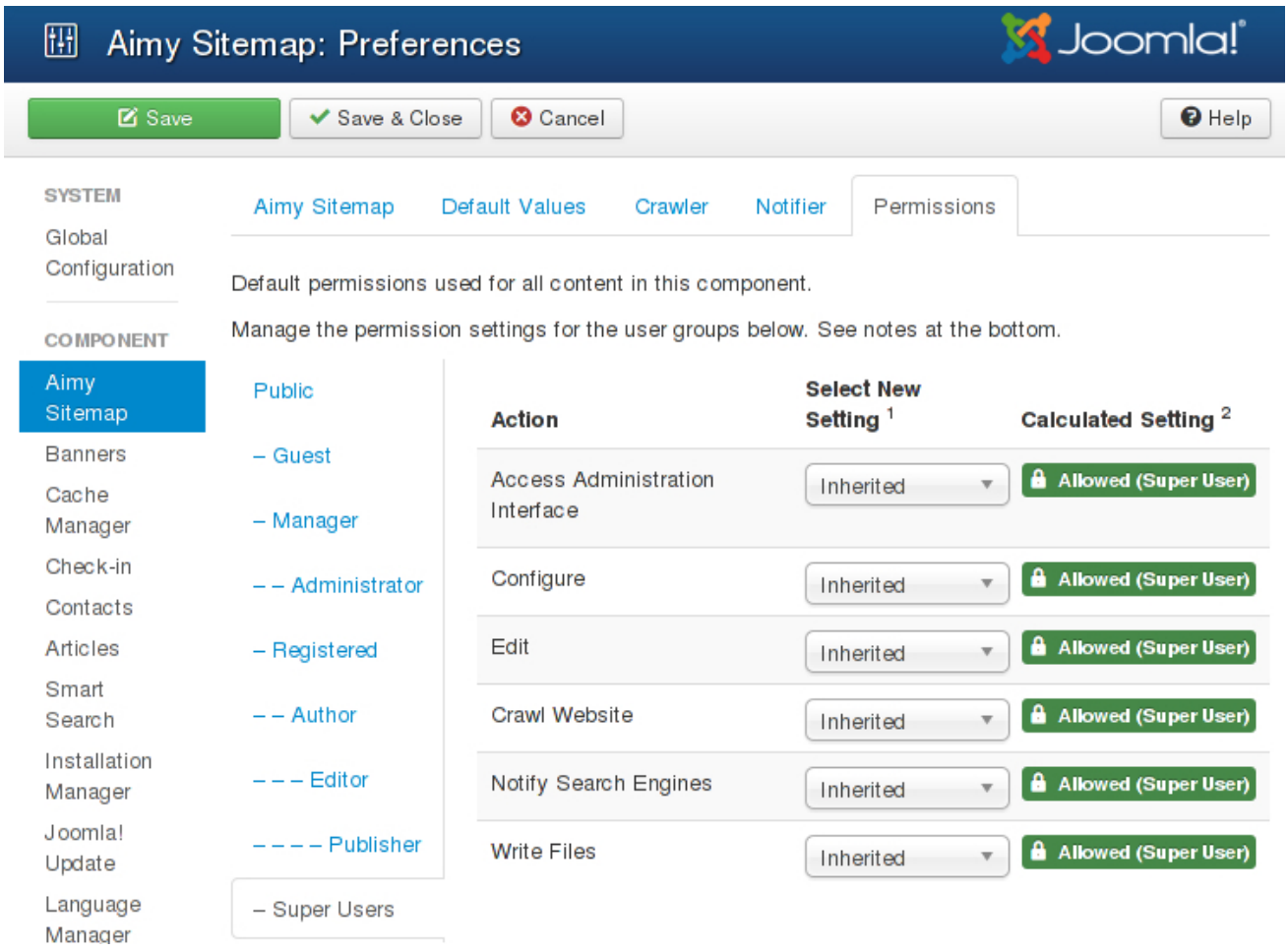
Write an updated version of your `sitemap.xml` file to disk.

Notify Search Engines

Notify all enabled search engines about the new version of your sitemap.

NOTE: Each action uses the options specified for the respective task so no additional configuration is necessary.

6.1.6 Tab: Permissions



Manage the permission settings for different user groups.

Aimy Sitemap allows to set permissions for the following actions:

Access Administration Interface

allows users to view Aimy Sitemap's administration interface.

Configure

allows users to view and change Aimy Sitemap's configuration.

Edit

allows users to edit URLs on the "Manage URLs" page and change their state.

Crawl Website

allows users to use Aimy Sitemap's crawler on your website and, as a result, update the set of URLs.

Notify Search Engines

allows users to notify the configured search engines about changes to your sitemap.

Write Files

allows users to initially write or update the configured **sitemap file** or the **robots.txt file** on your webspace.

For details, have a look at the official Joomla! Access Control List Tutorial:

`http://docs.joomla.org/J3.x:Access_Control_List_Tutorial`.

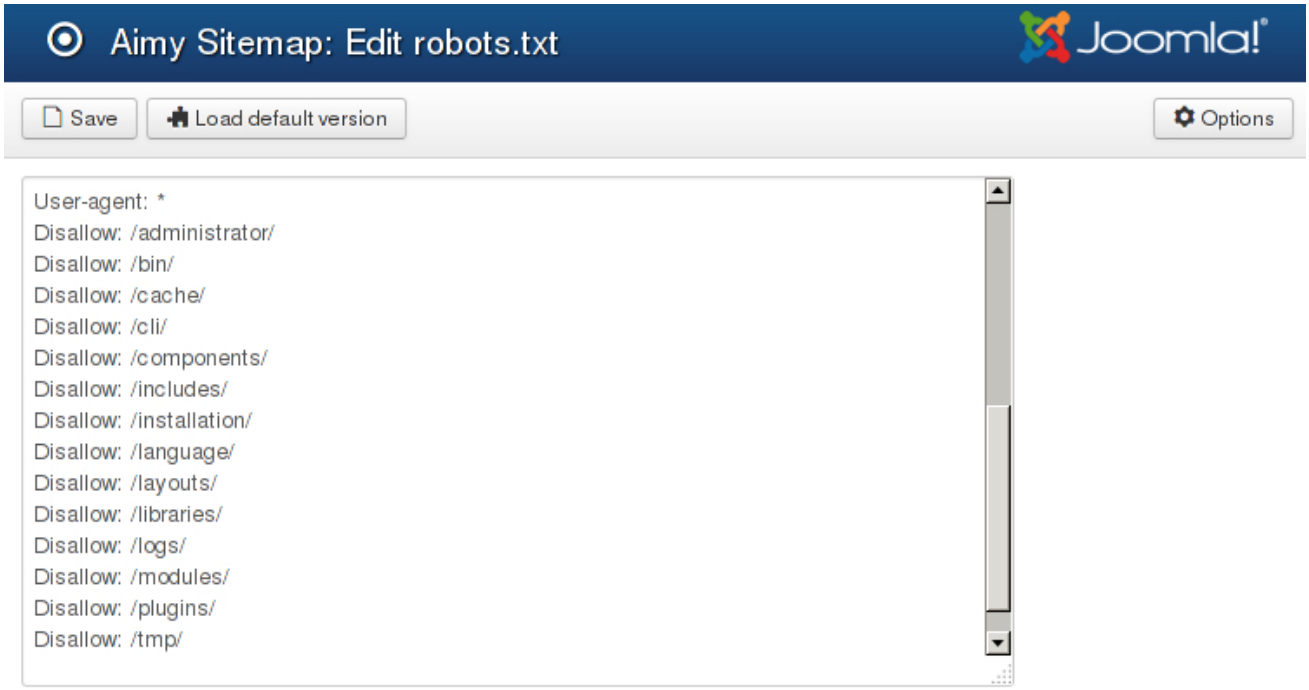
7 The robots.txt File

Aimy Sitemap comes with a simple `robots.txt` editor that allows you to set up your ruleset directly from Joomla!'s backend. You may use those rules to give instructions to Aimy Sitemap's crawler as well.

7.1 Editing

To edit your `robots.txt`, choose "*Edit robots.txt*" from Aimy Sitemap's menu. Your current file will automatically be loaded if present in the root directory of your Joomla! installation - otherwise a default version will be loaded. To write your changes to disk, click "*Save*".

At any time, you can click "*Load default version*" to load the default `robots.txt` file that comes with Aimy Sitemap.



7.2 How To Set Rules For Aimy Sitemap's Crawler

Aimy Sitemap will not crawl URLs that are either forbidden for any bot ("`User-agent: *`") or especially for Aimy Sitemap's crawler, called `AimySitemapCrawler`.

Use the following syntax in that case:

```
User-agent: AimySitemapCrawler
Disallow: /images/my-document.pdf
Disallow: /images/my-image.jpg
```

7.3 Further Information

For further information on the `robots.txt` file, visit `http://www.robotstxt.org/`.

8 Crawling Your Joomla! Website

If you have just installed Aimy Sitemap on your Joomla! website, you will not see any entries in the **Manage URLs** view yet. You have to crawl your website first. To do so, choose "*Components*" → "*Aimy Sitemap*" → "*Crawl Website*" and click the "*Start Crawling*" button in the toolbar.

After crawling has finished, you can view all URLs included in your index by clicking "*Manage Your Sitemap Now*" or by choosing "*Manage URLs*" from Aimy Sitemap's menu.

If you have finished looking through the list and configuring your individual URLs settings, click "*Write Sitemap*" to write the XML sitemap to your webserver's disk.

9 Managing Your Index and Sitemap

The "*Manage URLs*" view allows you to have a look at the index generated by Aimy Sitemap's crawler and manage your sitemap by setting attributes of URLs and by selecting which URLs should be included in your final sitemap.

If you discover items you do not want to display in the sitemap after crawling, **disable** the items in the list. Click the button in the "*State*" column for any single file or use the checkboxes and change the state of multiple files at once.

Next time you click the "*Write Sitemap*" button, all active URLs will be written to the sitemap file.

9.1 Setting Attributes

You may set the following attributes individually per document:

- State / Include in Sitemap

  Set whether the document should included in your sitemap.

- Priority

  Set the priority of the document for your website's content, ranging from 0.1 (less important) to 1.0 (very important).

- Change Frequency

  Set how often the document is regularly changed.

- Document Language (PRO feature)

  This attribute affects the HTML view of the sitemap for multi-language websites. See *Generating an HTML Sitemap*.

- Title (PRO feature)

  The title is relevant for the HTML view of the sitemap.

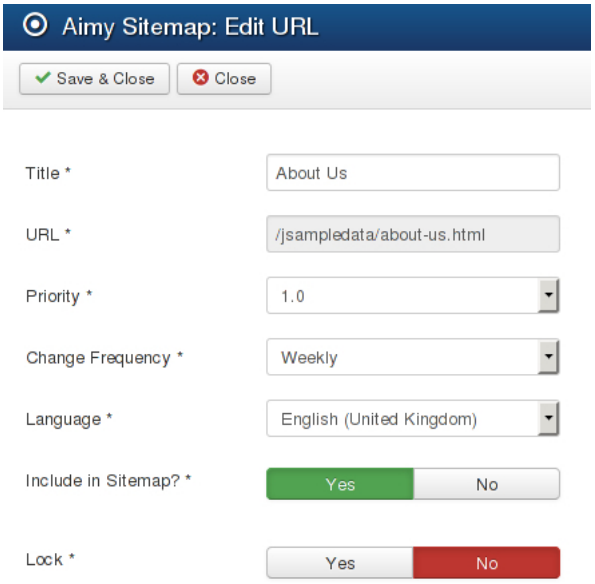
- Lock

  Set whether the document's language and title data should be updated by the crawler.

Note: If you have manually set either language or title, be sure to set the lock attribute to "*Yes*" in order to disable automatic updates.

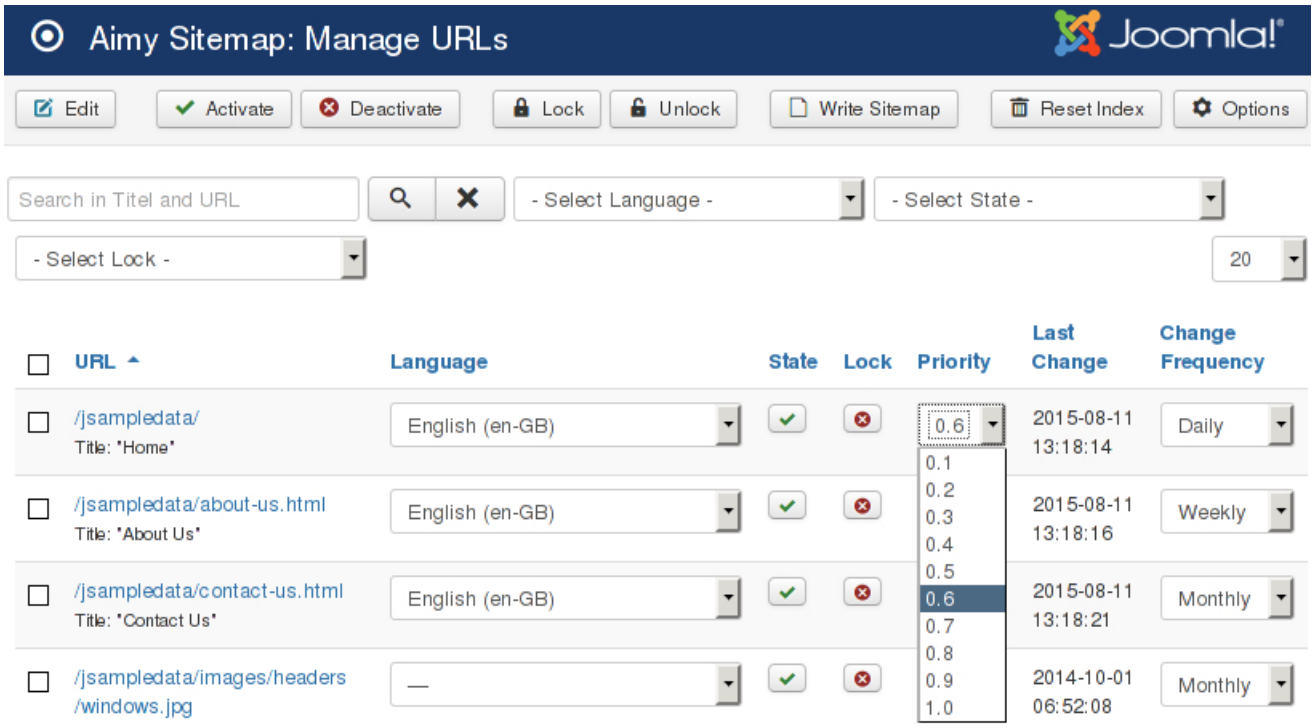
The settings can be done either in

1. a **dedicated edit view** for each URL reached by clicking on the URL or selecting the URLs checkbox and clicking the *Edit* button in the toolbar



2. **quick edit mode** that allows you to set your URLs' attributes directly from the URL list (requires *Quick Edit* to be enabled in *Options*).

   After you changed a value it will be saved automatically in the background. Success or failure is indicated by the background color of the changed field, if it turns green your change has been saved correctly, if it turns red it could not be stored - use the per URL edit-view in this case.



9.2 Searching, Filtering and Sorting

Aimy Sitemap allows you to search and filter your URLs to select just the set you want to work on. The following tools are available and combinable:

Search in Title and URL

Type your term into the **search field** and either press ENTER or click the magnifier button next to the field to start the search. The set of shown URLs will be updated according to the search results.

To reset your search, click the "X" button next to the **search field**.

Filter by Language (PRO feature)

This filter allows you to select URLs from their document's language. Your selection will take effect automatically.

To reset your selection, select "-Select Language-" from the drop-down list.

Filter by State

This filter allows you to select whether to show "activated", "deactivated" or "all" URLs. Your selection will take effect automatically.

To reset your selection, select "All" from the drop-down list.

Filter by Lock (PRO feature)

This filter allows you to select whether to show "locked", "unlocked" or "all" URLs. Your selection will take effect automatically.

To reset your selection, select "All" from the drop-down list.

Sort by Column

To sort your URLs by value of one of the columns, click the column's heading. By clicking the heading once, the URLs will be sorted ascending, a second click sorts descending. A small caret will be displayed right to the column's name indicating the sorting direction.

The "*URL*" and "*Change Frequency*" columns are sorted alphabetically, "*Priority*" is sorted numerically and "*Last Change*" is sorted by date.

Pagination

For convenience, the URLs of your index will be displayed grouped by virtual pages of a customizable amount of URLs each that you can navigate through using a toolbar below the list of URLs.

You can set a custom amount of URLs per page by selecting your desired number from the drop-down box in the upper right just above the list of URLs.

9.3 When to Re-Crawl Your Website?

If you have

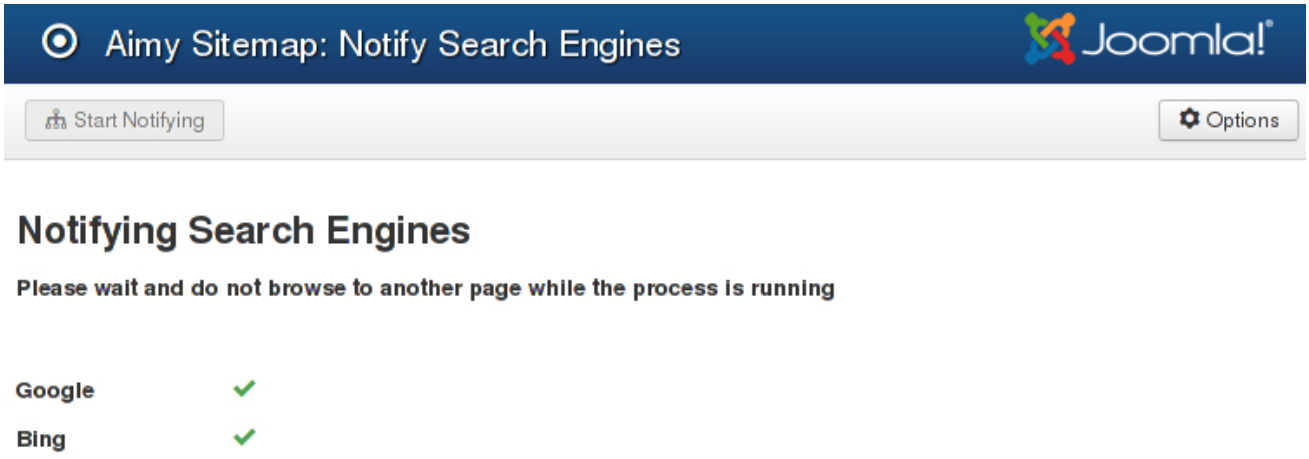
- changed your content (renamed, added or deleted articles) or
- changed crawler configuration (e.g. exclude patterns)

crawl your Joomla! website again, view the result and click "*Write Sitemap*".

10 Notifying Search Engines

You can notify search engines about your new sitemap or about changes. Choose "*Notify Search Engines*" from the menu and click "*Start Notifying*" on the toolbar.

This will send a ping request to the search engines you activated in Aimy Sitemap's configuration.

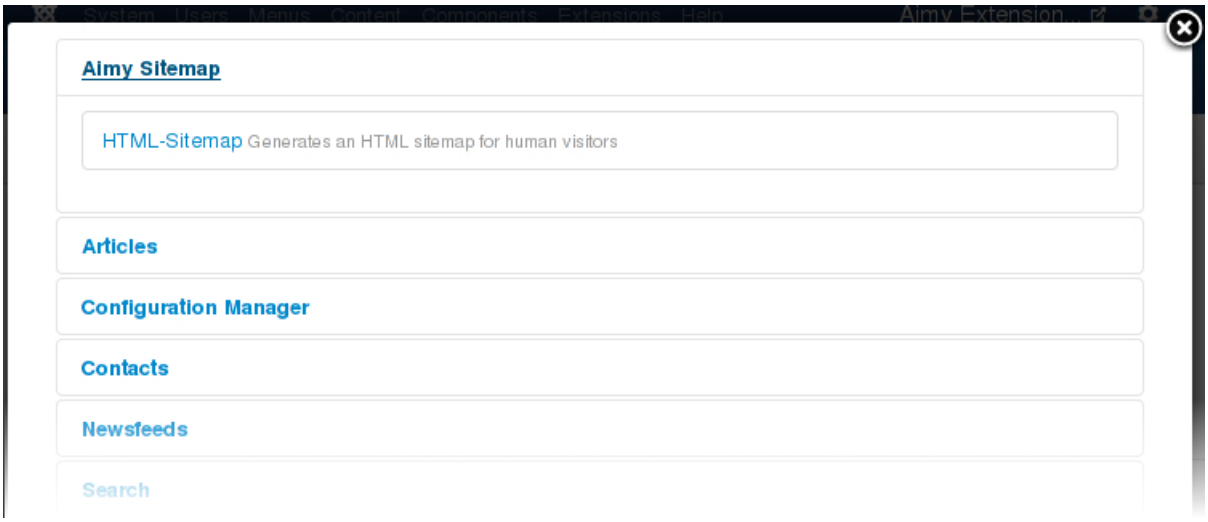


11 Generating an HTML Sitemap

Aimy Sitemap provides different HTML views of your sitemap. If you want to display an HTML sitemap on your website, add it as a menu item by clicking:

"*Menus*" → "*YOUR MENU*" → "*Add New Menu Item*"

As *Menu Item Type* choose "*Aimy Sitemap*" → "*HTML-Sitemap*".



11.1 Options

Aimy Sitemap's HTML views allow you to set the following options beside the standard Joomla! options for a menu item:

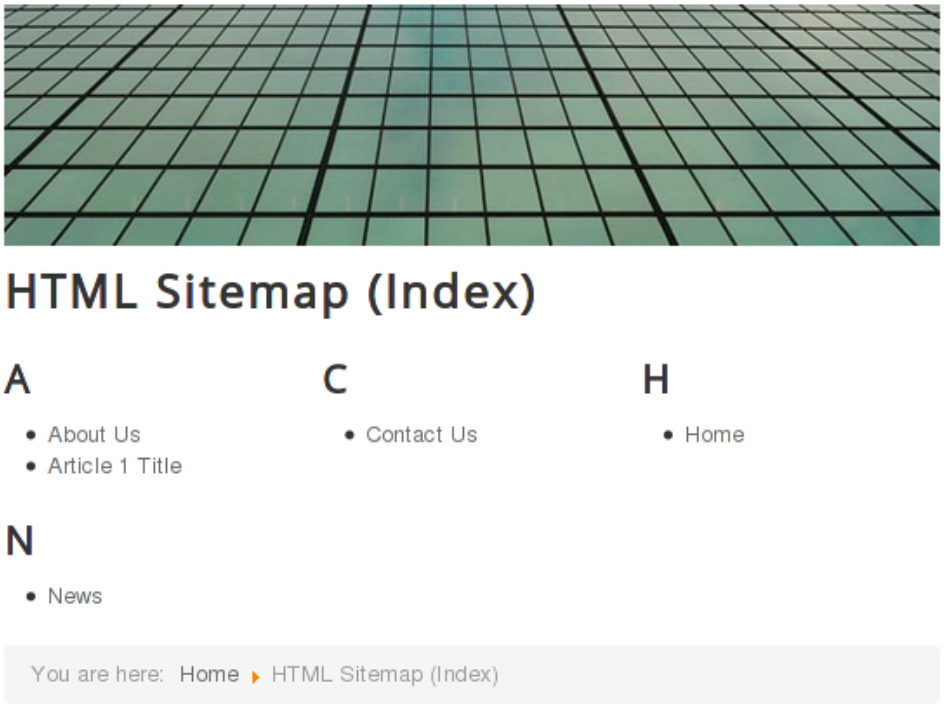
Variant

List

All URLs are displayed as a simple top-to-bottom list, sorted alphabetical by their title.

Index

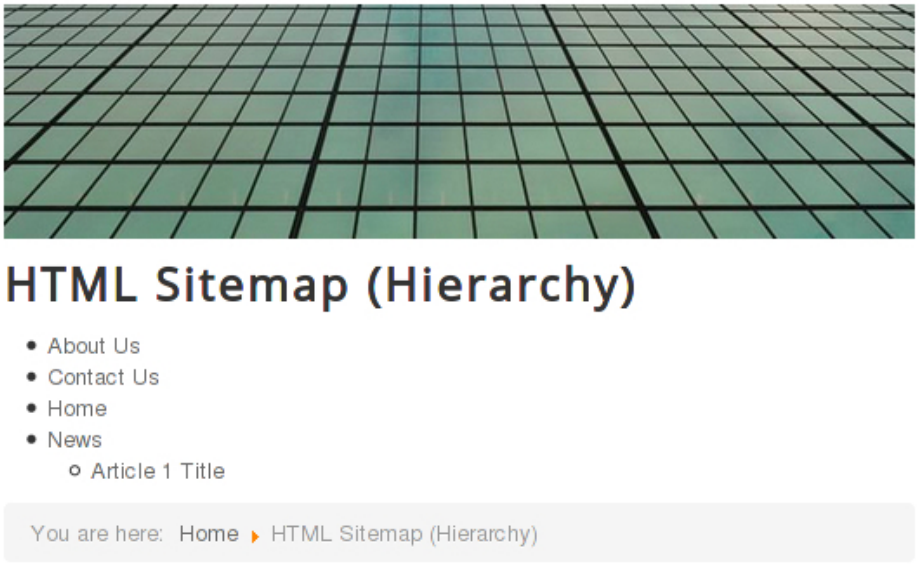
All URLs are displayed in blocks named after their first letter. Both blocks and the contained links are sorted alphabetical by title.



Hierarchy (PRO feature)

URLs are displayed hierarchically, with each collection being sorted alphabetically.

NOTE: This feature requires that SEF URLs are enabled.



Notes On Styling

You can use CSS to style the list elements.

Levels

Each `ul` element provides a CSS class that reflects its level. The class is named `aimysitemap-lvl-$L`, with `$L` being the numeric level, i.e. `aimysitemap-lvl-2`.

The top-level `ul` element has a level of 1 (hierarchy root).

Languages

Each link that has an assigned language does provide this information as a CSS class as well. The class is named `aimysitemap-lang-$L`, with `$L` being the language code in lowercase letters, i.e. `aimysitemap-lang-de-de` for a link to a German document and `aimysitemap-lang-en-gb` for an English one.

Prevent Duplicate Titles

If this option is set to *Yes*, all URLs will be merged by their title so that no title can be present on your HTML sitemap more than once.

NOTE: If this option is on, the *shortest URL* of the set of URLs sharing the same title will be selected.

BACKGROUND: In some cases Joomla! generates multiple URLs for a resource. As a result, Aimy Sitemap's crawler adds these resources to the index multiple times as well. That's technically correct. As these URL will likely have the same title, it may seem wrong to your human visitors to show them more than once in your HTML sitemap. That's what this option is for.

Filter by Language (PRO feature)

If this option is set to *Yes*, Aimy Sitemap will only show entries that match the frontend language. You see - and may change - the language of the entries in *Manage URLs*.

The titles for the sitemap are extracted from the page's title tag provided by Joomla! automatically. In the PRO version of Aimy Sitemap you may set the title manually as well (see *Managing your Index and Sitemap → Setting Attributes*).

12 Periodic Crawling (PRO feature)

Aimy Sitemap PRO supports crawling your website automatically. After such a **periodic crawl** your *Index* is updated. Additionally the `sitemap.xml` file may be updated and search engines notified if new or updated content is found during the crawl. These actions depend on your settings in the *options*.

Detailed information about the last periodic crawl is stored and can at any later time be viewed in Joomla!'s administration backend.

12.1 Obtaining the Script's Path

The Aimy Sitemap extension includes a command line script named `crawl-website.php` that allows periodical crawling. In order to execute this script, you first have to know its absolute path.

To look the path up, choose "Components" → "Aimy Sitemap" → "Periodic Crawl" in your Joomla! administration backend. The absolute path on your server is shown under "*Help*":

Aimy Sitemap: Periodic Crawling
Joomla!
Show Report
Show Output
Options
Last Periodic Crawl
There has been no periodic crawl registered until now.
Help
Path to the script on your server:
/aimy-demo/administrator/components/com_aimysitemap/cli/crawl-website.php
For further information on how to configure periodic crawls have a look at the user manual or our blog posting on how to set up Microsoft's Task Scheduler.

12.2 Scheduling Automatic Crawls

In order to schedule an automatic crawl of the sitemap generator, a **system service** is required. On a Unix or Unix-like system (i.e. Linux, FreeBSD, Mac OS X, Solaris) **cron** is usually used for such tasks while on a Microsoft Windows server, **Task Scheduler** is available. Most hosters provide an interface to schedule jobs for periodical execution as well.

12.2.1 Setting Up a Cron-Job (Unix)

If you have a Unix or Unix-like server with shell access, log into your server using the user account that should be running the automatic crawl periodically. This may be the same user your Joomla! website is running as.

First, determine the path of your PHP interpreter. If you do not already know where the binary is located, run the following command to find out:

```
which php5 || which php
```

Then, add a new cron-job running the following command:

```
crontab -e
```

Your default editor will open up, allowing you to enter a new cron-job. Each cron-job consists of the following six required fields: *minute*, *hour*, *day of month*, *month*, *day of week*, *command*. The special value of "*" can be used as placeholder for any valid value in one of the first five fields, i.e. any hour or any day of the week.

Set up your cron-job to be run at your desired time and be sure to use the full path to your PHP binary followed by the full script path as a *command*.

To suppress all non-error output during execution, you can pass either `-q` or `-quiet` as an argument.

Save your changes and exit your editor afterwards.

Examples:

To run your periodic crawls each night at 23:55 you would enter:

```
55 23 * * * $full_php_path $full_command_path
```

NOTE: Replace `$full_php_path` with the full path of your PHP interpreter (i.e. `/usr/bin/php`) and `$full_command_path` with the path you previously looked up in Aimy Sitemap's "*Periodic Crawl*" view.

To run your periodic crawls each Sunday at noon, suppressing all non-error output, you would enter:

```
00 12 * * sun $full_php_path $full_command_path --quiet
```

12.2.2 Setting Up a New Task (Windows)

On a Microsoft Windows server, **Task Scheduler** can be used to schedule automatic website crawls periodically. To start the utility, click "Start" → "Accessories" → "System Tools" → "Task Scheduler".

Create a new "Basic Task" and enter your desired configuration.

When asked what *Action* should be performed, choose "*Start a program*". Choose your PHP interpreter as a program and enter the full path to the `crawl-website.php` script in the "*Add arguments*" field below.

To suppress all non-error output, add `-quiet` as an argument (after the script's path) in the "*Add arguments*" field.

Verify your settings in the next mask and save the task by clicking "*Finish*".

For an in-depth step-by-step howto on how to set up a task for Aimy Sitemap's periodic crawling script using Microsoft's Task Scheduler have a look at our website.

12.2.3 Using Your Hosters Interface

Most hosters provide an interface to the periodic command execution system services in their management frontends. First, verify that your hoster provides such an interface for your website and have a look at your hoster's documentation.

When asked for it, choose the highest version of PHP available (i.e. 5.6) as an interpreter and use the full path of the `crawl-website.php` script as periodic task.

By default, PHP processes started from the command line do not have a time limit set. However, if your hoster has overwritten this default setting (`max_execution_time`), be sure either restore the default (0) or at least set the highest limit available for the directive (>60) to decrease the probability that the `crawl-website.php` script will get aborted by your system.

12.3 Crawling Information

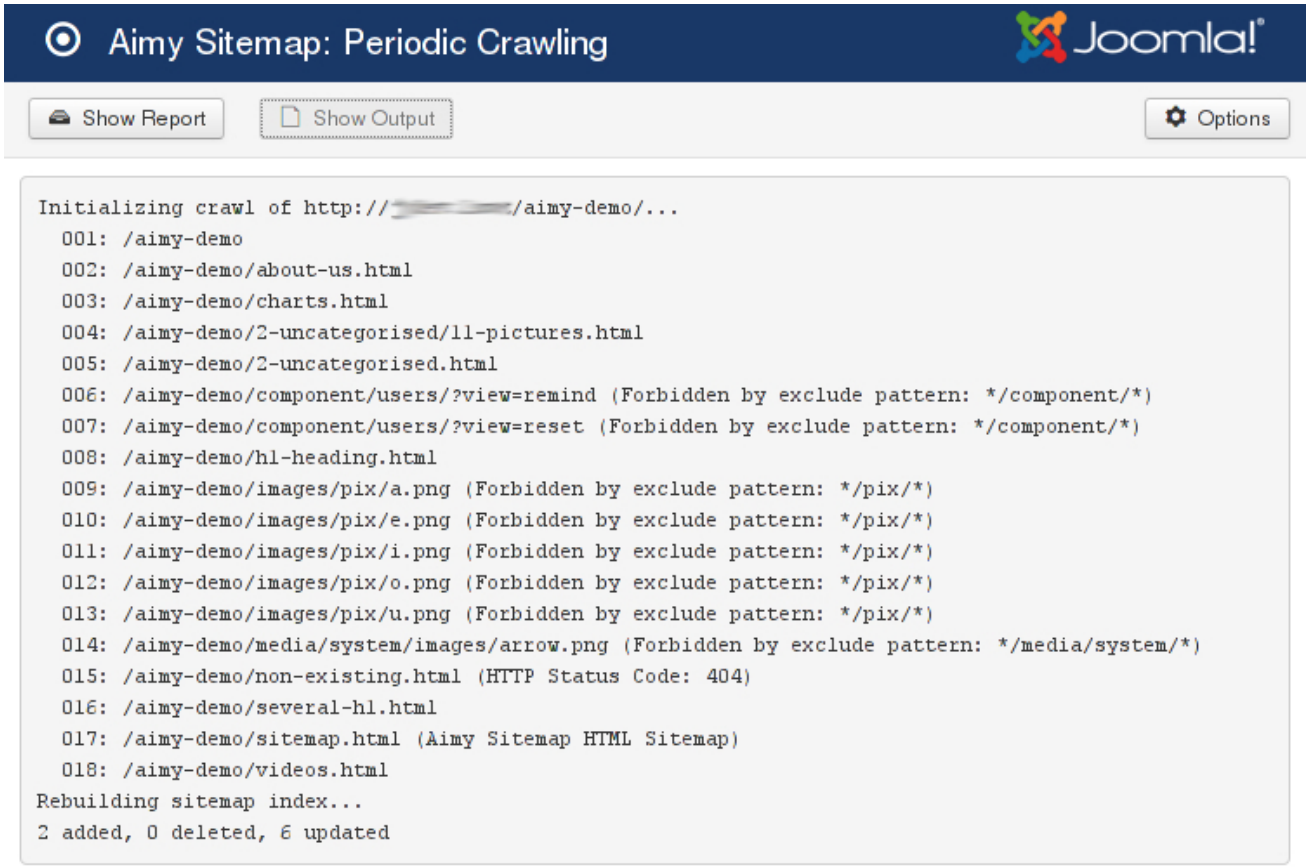
Information about the last periodic crawl, including a detailed report of changes found and the whole command line output, can be found within your website's Joomla! administration backend: "Components" → "Aimy Sitemap" → "Periodic Crawl".

The default view provides a short report on the last crawl, along with all errors that may have occurred during the crawl and a short help:

Aimy Sitemap: Periodic Crawling
Joomla!
Show Report
Show Output
Options
Last Periodic Crawl
The last periodic crawl started at 2015-06-24 09:25:14 and took 0 hour(s), 0 minute(s) and 14 second(s).
18 URLs have been crawled and 8 modifications have been done to your index: 2 URLs added, 0 deleted and 6 updated.
Errors
015: /aimy-demo/non-existing.html (HTTP Status Code: 404)
Help
Path to the script on your server:
aimy-demo/administrator/components/com_aimysitemap/cli/crawl-website.php
For further information on how to configure periodic crawls have a look at the user manual or our blog posting on how to set up Microsoft's Task Scheduler.

To view the detailed report on all changes done to your index, click the "*Show Report*" button:

To view the full output of the last periodic crawl, click the "*Show Output*" button:


Aimy Sitemap: Periodic Crawling
Joomla!
Show Report
Show Output
Options
Initializing crawl of http://[redacted]/aimy-demo/...
001: /aimy-demo
002: /aimy-demo/about-us.html
003: /aimy-demo/charts.html
004: /aimy-demo/2-uncategorised/11-pictures.html
005: /aimy-demo/2-uncategorised.html
006: /aimy-demo/component/users/?view=remind (Forbidden by exclude pattern: */component/*)
007: /aimy-demo/component/users/?view=reset (Forbidden by exclude pattern: */component/*)
008: /aimy-demo/h1-heading.html
009: /aimy-demo/images/pix/a.png (Forbidden by exclude pattern: */pix/*)
010: /aimy-demo/images/pix/e.png (Forbidden by exclude pattern: */pix/*)
011: /aimy-demo/images/pix/i.png (Forbidden by exclude pattern: */pix/*)
012: /aimy-demo/images/pix/o.png (Forbidden by exclude pattern: */pix/*)
013: /aimy-demo/images/pix/u.png (Forbidden by exclude pattern: */pix/*)
014: /aimy-demo/media/system/images/arrow.png (Forbidden by exclude pattern: */media/system/*)
015: /aimy-demo/non-existing.html (HTTP Status Code: 404)
016: /aimy-demo/several-h1.html
017: /aimy-demo/sitemap.html (Aimy Sitemap HTML Sitemap)
018: /aimy-demo/videos.html
Rebuilding sitemap index...
2 added, 0 deleted, 6 updated

NOTE: If you passed "`-quiet`" as an argument to the "`crawl-website.php`" script, no non-error output is sent to the command line. For convenience, the output that would have been generated is stored and provided for review nevertheless.

13 Special Cases

13.1 Crawling HTTPS-only Websites (PRO feature)

Aimy Sitemap PRO supports crawling of HTTPS-only websites, this is, of websites that cannot be accessed unencrypted using HTTP.

The only requirement is that your **PHP installation** does support SSL.

To find out whether your PHP installation supports SSL, log into your Joomla! backend and choose "*System*" → "*System Information*" and click the "*PHP Information*" tab. In the first table, look for "*Registered Stream Socket Transports*".

If the "**ssl**" transport is listed in this set of stream transports, Aimy Sitemap's crawler will be able to crawl your website using https.

13.2 Joomla! Installation in Subdirectory

If you have installed Joomla! to a **subdirectory** and want to use the `sitemap.xml` file Aimy Sitemap generates as a top-level sitemap nevertheless, you cannot directly set the path in Aimy Sitemap's options. That is, because for security reasons Aimy Sitemap will non write a sitemap to a file outside of Joomla!'s root directory.

However, there are a couple of ways to use the sitemap located within the Joomla! subdirectory as a top-level sitemap so that search engines will find and use it:

13.2.1 Rewrite Rule

You can set a rewrite rule for your webserver so that it "virtually" serves `/sitemap.xml` using the one created in your subdirectory. This approach is quite common, for example Joomla! itself uses rewrite rules for search engine friendly URLs.

The following rule is specific for *Apache*. If you use a different webserver for your site, have a look at the software's manual on how to write a similar rule.

Add the following lines to your **top-level** `.htaccess` file or create a new one if it does not already exists:

```
RewriteEngine on
RewriteCond %{REQUEST_FILENAME} !-f
RewriteRule ^sitemap\.xml$ /SUBDIR/sitemap.xml [L]
```

Just replace `SUBDIR` with the name of your subdirectory.

13.2.2 Sitemap Index File

Another way to put your subdirectory sitemap to use is to create a *Sitemap Index*. That's a simple XML file telling a search engine where to look for further sitemaps.

Here's an example for a top-level `sitemap.xml` file that references your real sitemap in your subdirectory:

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <sitemap>
    <loc>http://YOURDOMAIN/SUBDIR/sitemap.xml</loc>
  </sitemap>
</sitemapindex>
```

Just replace YOURDOMAIN and SUBDIR with the appropriate values for your setup.

For more information have a look at `http://www.sitemaps.org/protocol.html#index`.

13.2.3 Robots.txt Directive

You may use a `robots.txt` directive to give search engines a hint on where to look for your sitemap as well. Please note that this directive is an extension of the standard robots.txt specification that may not be supported by all search engines.

Add the following line to your `robots.txt` file:

```
Sitemap: /SUBDIR/sitemap.xml
```

13.3 Using a Canonical URL (PRO feature)

If your Joomla! website can be addressed using more than one domain name, one way to prevent search engines interpreting it as having *duplicate content* is to set a *canonical link tag* on your pages.

Our free plugin **Aimy Canonical** allows you to set a *unique domain name* and your preferred protocol (`http` or `https`) that is used to create the *canonical link tag* for all of your website's pages.

If Aimy Canonical is installed and enabled, Aimy Sitemap will read the settings as well and keep your sitemap's URLs in sync with it.

Feel free to download the Aimy Canonical Joomla! plugin on our website:

`http://www.aimy-extensions.com/joomla/canonical.html`

14 Notes and Limitations

- The "Manage URLs" backend view may not represent the current **sitemap on disk**: keep in mind to click "*Write Sitemap*" to apply changes, like updating frequency, priority or state.
- If you changed your **preferences for the crawler**, for example the inclusion of images, crawl again to update your index.
- Any change of the **default values** will –by definition– apply for new entries only.

  If you like new default values to apply, click the "*Reset Index*" button in "Manage URLs" and re-crawl your website. Your new default values will then be applied to all found URLs.
- The **sitemap file** will be deleted when Aimy Sitemap is uninstalled as well.
- Only URLs of up to 767 bytes in length are indexed (URL encoded).

- Websites which are put into **offline mode** cannot be crawled as expected. That's because in offline mode there are no links to your content that Aimy Sitemap's crawler could extract and analyze.
- Documents which require **HTTP authentication** ("HTTP Auth") cannot be crawled. Including those in your sitemap wouldn't be useful, because search engines won't be able to index them as well.

15 Debugging and Errors

If you have problems crawling your Joomla! website, this may have a couple of reasons. If none of our hints in the backend or manual helps to solve the problem, you can turn on **debugging mode** to get some more details.

Go to you Joomla! configuration in *System* → *Global configuration* → Tab *System* and set *Debug System* to "Yes".

If you start a crawl afterwards it will be logged in a file named "`aimysitemap.php`" in the "`logs`" directory of your Joomla! webroot. Go through the file to find hints on the issue or send the file to Aimy Extensions' support team for review and help.

Note: If debugging mode is on, no index will be generated.

Additionally there may be common problems depending on your setup. In the FAQ on our website you find some hints that may help you as well:

`http://www.aimy-extensions.com/joomla/sitemap.html#faq`

16 Copyright & Trademark Notice

The Joomla!® name and logo are trademarks of Open Source Matters, Inc. in the United States and other countries.

Mentioned hard- and software as well as companies may be trademarks of their respective owners. Use of a term in this manual should not be regarded as affecting the validity of any trademark or service mark. A missing annotation of the trademark may not lead to the assumption that no trademark is claimed and may thus be used freely.